Ab-initio nuclear magnetic resonance file format v 1.0

Tim Green and Jonathan Yates
Oxford Materials

February 3, 2014

1 Aims

- Simple is beautiful.
- Provide a complete archival format for a first principles calculation of NMR parameters - e.g. for providing as SI to an article or for storage in an archival database
- Common format across first-principles software.
- Provide a platform on which to provide interface to spin-simulation codes, visualisation etc
- Designed to be easily accessible to researchers for further processing
- Easy to make a parser in any standard language.

2 Requirements

- Allow standard command-line tools to be used to extract information, e.g. grep, awk.
- Allow unrecognised parts of the file to be ignored, allowing easy extension with other atomic properties
- Not to specify redundant or arbitrary information, to reduce maintenance and chance of contradiction.
- Properties specified as full cartesian tensors in the same basis as the atomic positions in well-defined units. Extraction of further information, e.g. eigenvalues or isotropic values, is up to further processing.

- Crystal information of lattice and atoms provided for visualisation, calculation of geometric properties, etc.
- Not to use XML on the basis that bringing in parsing libraries can be cumbersome in many languages and such complexity is unnecessary.
- Numerical accuracy.

3 Limitations

- At the present time only parameters related to NMR of diamagnetic materials have been specified. The extension to include parameters related to paramagnetic materials should be straight forward.
- The present standard doesn't include definitions for providing EPR parameters. This could be incorporated in a future backwards-compatable standard.
- While the format is designed for processing and archiving, it's not designed for easy querying in a database form. It's assumed a database will perform further transformations into its own desired internal format for indexing.

4 Specification

The file is essentially a series of blocks consisting of rows constaining whitespace-delimited records, with datatype distinguished by tags in the first column. To allow easier recognition by parsers, the file must start with "#$magres-abinitio-v1.0", the version number 1.0 referring to this standard. Minor version number increases after the period will aim to not break backwards compatability with existing parsers whereas major version number increases before the period may.

4.1 Blocks

Blocks are wrapped with tags:

```
[block]
  ... data ...
[/block]
```

They cannot be nested or overlap. Unrecognised blocks should be ignored, so as to allow other data to be put in the same file. Blocks discussed in this specification are

- `[atoms]` - Atomic coordinates and optional lattice. Must be first block for easy parsing.

- `[magres]` - Calculated magnetic resonance parameters
- `[calculation]` - Calculation metadata such as code version and input parameters

The `[atoms]` block should appear before `[magres]` for ease of parsing. The feature of ignoring unrecgonised blocks allows codes to include extra blocks of text, such as old magnetic resonance file types that have to be retained for backwards compatability with existing parsers. Input files could also be kept in extra unrecognised blocks.

4.2 Characters and numerical representation

This file format is based on 8-bit ASCII characters but only characters 0-127 should be used outside of comments.

Numbers should be written in scientific notation with double precision (i.e. typically 16 significant figures). Numbers should be written as to be easily parsable by standard string routines (e.g. Fortran, C, Python, Perl, etc.)

4.3 Records

A record is a line in the file. Whitespace (characters recognised by \s in a regex) is ignored at the start and end of a line. One or more whitespace characters separate two entries in a record. Each record has a tag, which is the first entry.

```
lattice 6.0 0.0 0.0 0.0 6.0 0.0 0.0 0.0 6.0
```

In the above record, "lattice" is the tag.

4.4 Comments

Any characters following a # symbol in a line should be ignored.

4.5 Defined tags

4.5.1 Units

Although this specifiation document will specify the units for each parameter, it's preferable to explicitly define the units used for each kind of tag for discoverability and so as to allow some flexibility and allow parsers to hard-fail when unrecognised units are used. The tag for units is:

```
units <tag> <units>
```

where the units is some case-sensitive string that would make sense to a parser and human, e.g "au" for when atomic units have been used. The strings should be strictly compared, i.e. "J.mol-1" is not the same as "Jmol^1".

The `units` tag must appear once or not at all before the first instance of the tag it is describing. Parsers should warn if they do not recognise units. If they have to process quantities with units they don't recognise they should fail.

4.5.2 `atoms` block

The crystal lattice can be given as the $\mathbf{a}, \mathbf{b}, \mathbf{c}$ unit cell vectors. This is an optional element, however if it is not present it can be assumed that the structure is non-periodic. We will standardize on Angstrom as the output unit.

```
units lattice Angstrom
lattice a1 a2 a3 b1 b2 b3 c1 c2 c3
```

The atomic positions (for all the atoms in the full crystallographic unit cell) are given in cartesian coordinates, standardized to Angstrom as the output unit. These must be present in a file before any magres parameters.

```
units atom Angstrom
atom <atom type> <species label> <number in species label> x y z
```

The atom type should be an element symbol (no isotopes, i.e. H rather than D) and the species label can be any string, e.g. the crystallographic label. If there is no label the element symbol should be given again. Atoms can be specified out of order, but the numbers in the label should be in order.

If symmetry operators are known, you can specify them with the `symmetry` tag:

```
symmetry <symmetry string>
```

where the symmetry string is in x,y,z notation. As it's undesirable to require basic scripts to know and be able to apply symmetry operators, all non-unique atoms should still be included. An example set of symmetry operations for an alanine crystal structure would be

```
symmetry x,y,z
symmetry x+1/2,-y+1/2,-z
symmetry -x,y+1/2,-z+1/2
symmetry -x+1/2,-y,z+1/2
```

4.5.3 `magres` block

`ms` - Magnetic shielding

The magnetic shielding tensor is defined as the ratio between this induced field, and the external applied field.

$$\mathbf{B}_{\text{in}}(\mathbf{R}) = -\sigma(\mathbf{R})\mathbf{B}_{\text{ext}}(\mathbf{R}) \tag{1}$$

$\sigma(\mathbf{R})$ is a dimensionless quantity which is reported in parts per million (ppm). It is a rank-2 tensor defined for each atomic site and specified by 9 independent Cartesian components. Note that it is not a symmetric tensor but can be written as the sum of symmetric and antisymmetric tensors.

```
units ms ppm
ms <species label> <number in species label> 11 12 13 21 22 23 31 32 33
```

Where the `XY` indices label the components of the tensor like

$$\begin{pmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{pmatrix} \tag{2}$$

`efg` - Electric field gradient

The Electric Field Gradient (EFG) is a rank-2, symmetric, traceless tensor $V(\mathbf{R})$ given by

$$V(\mathbf{R}) = \frac{\partial \mathbf{E}_\alpha(\mathbf{R})}{\partial r_\beta} - \frac{1}{3}\delta_{\alpha\beta}\sum_\gamma \frac{\partial \mathbf{E}_\gamma(\mathbf{R})}{\partial r_\gamma} \tag{3}$$

where α, β, γ index the Cartesian coordinates x,y,z and $\mathbf{E}(\mathbf{R})$ is the local electric field at the position $\mathbf{R}$. The EFG is recorded in atomic units. Note that 1a.u. of EFG is approximately $9.717 \times 10^{21} V m^{-2}$.

```
units efg au
efg <species label> <number in species label> 11 12 13 21 22 23 31 32 33
```

A code might provide the EFG decomposed into theory-dependent contributions with additional tags (e.g. `efg_ions` for just the contribution from the atomic nuclei + cores).

`isc` - **Indirect spin-spin coupling (J-coupling)**

The indirect spin-spin coupling tensor is obtained from the magnetic field induced at nucleus K due to the perturbative effect of the magnetic moment, μ_L, of nucleus L,

$$\mathbf{B}_{\text{in}}(\mathbf{R}_L) = K_{KL} \cdot \mu_L \tag{4}$$

K_{KL} is known as the reduced spin coupling tensor. It is a rank-2 tensor, and is given in units of $10^{19}\text{T}^2\text{J}^{-1}$. The J-coupling tensor is obtained from this using $J_{KL} = 2\pi\hbar\gamma_K\gamma_L K_{KL}$ where γ_K is the gyromagnetic ratio of the nucleus K. Note that K (and hence J) is not symmetric but can be written as the sum of symmetric and antisymmetric tensors. The tensors K_{KL} and K_{LK} are different. The symmetric part is unchanged on exchanging sites K and L, but the antisymmetric part will change sign.

```
units isc 10^19.T^2.J^-1
isc <species label K> <number in species label K> <species label L>
  <number in species label L> 11 12 13 21 22 23 31 32 33
```

A code might provide the K-tensor decomposed into theory-dependent contributions with additional tags.

`sus` - **Macroscopic magnetic susceptibility**

The Macroscopic Magnetic Susceptibility of the sample. This is a single rank 2 tensor reported in units of $10^{-6}\text{cm}^3\text{mol}^{-1}$

```
units sus 10^-6.cm^3.mol^-1
sus 11 12 13 21 22 23 31 32 33
```

4.5.4 Calculation block

The exact content of this section is not precisely defined as applications are not clearly defined, however we suggest it contain sufficient information for a user to reproduce the calculation. For a calculation using a planewave pseudopotential code this might include

- date of calculation
- version of software
- pseudopotentials (files or generation strings)
- planewave cutoff energy
- kpoint sampling

Example:

```
calc_name                   ethanol-666
calc_comment                Ethanol in a 6x6x6 Angstrom cell
calc_code                   castep
calc_code_version           5.5
calc_cutoff_energy          600 eV
calc_kpoint_spacing         0.01 2.pi.Angstrom^-1
```

It might also, for example, include a list of hashes of the input files.

4.6 Examples

Attached file "ethanol-nmr.magres" is an ethanol shielding and efg calculation.

Attached file "ethanol-jc.magres" is an ethanol J-coupling calculation.

Attached file "alanine.magres" is an alanine shielding and efg calculation including symmetry operators.

4.7 JSON serialisation

We also define an optional JSON serialisation of the format for interchange between tools. An example is given at `https://github.com/tfgg/magres-format/blob/master/samples/alanine.magres.json` and the jsonschema definition is given at `https://github.com/tfgg/magres-format/blob/master/magres/schema/schema.py`. By convention this has the file extension .magres.json .